Workflow - Conversion of PDF/scanned images to Microsoft Word format using ABBYY Finereader

1. Import document into Finereader

Note:
For scanned PDFs, you can improve image quality using Image Editor. This must be done for each page before Step 2. Useful tools include Brightness and Contrast, Deskew, Correct Trapezoid Distortion, Straighten Text Lines, etc.

2. Define structure and recognize each page

Notes:
For simple layouts, you can do page analysis automatically to get faster results.
For tables, mark a table area and recognize it automatically before correcting.
Re-recognize modified areas such as tables.

3. Check for recognition issues in Finereader

- low-confidence characters
- spelling errors

4. Save Document / Save Selected Pages

Notes:
If some pages need to be skipped, either delete them from Pages panel or Shift-select pages you want to convert and then use Save Selected Pages command.

5. Check text recognition errors in Microsoft Word

Check and fix
- low-confidence characters (highlighted)
- spelling mistakes

6. Automatically fix / check for formatting issues in Microsoft Word

6.1. TransTools - Unbreaker
- Remove wrong paragraph breaks/line breaks

6.2. TransTools – Document Cleaner – Autoformat / Quick Actions
- Fix paragraphs with font formatting variations.
- Remove frames.
- Identify textboxes and selectively move them into main story
- Replace tab characters between words with spaces
- Convert wrong tables of contents to text
- Other changes

6.3. TransTools – Document Cleaner – Tag Cleaner
- Remove redundant tags.

7. Manually format Word document

- Set headers / footers
- Merge multi-page tables
- Fix incorrect numbering in numbered lists
- Check for merged paragraphs and split them if appropriate
- Put text back into columns where appropriate
- Check bold, italic, underline against PDF
- Check text color and background colors
- Check paragraph alignment, indentation, spacing

8. Handle images containing tags

Option 1: Overlay textboxes on top of images.
Option 2: Translate text in TransPDF or PDF editor, correct, paste into final Word document as an image.
Option 3: Overlay text areas over image area in Finereader, translate in separate Word document, format, paste into final document as an image.

9. Translate

- Create to-do list for post-checking: checking of textboxes, updating tables of contents, segmentation on tab characters.
- Import, check for tags and incorrect segmentation, fix and re-import if appropriate

10. Pre-flight check of translated document

- Visual check against PDF
- Resize textboxes and table rows with partially invisible text
- Update tables of contents, etc.